

CRITICAL NOTICE

JUDEA PEARL

Causality: Models, Reasoning, and Inference

Cambridge: Cambridge University Press, 2000, £25

ISBN: 0521773628

Donald Gillies

Department of Philosophy

King's College London

Judea Pearl's new book contains both a remarkable array of new mathematical techniques concerned with causality, and also much philosophical discussion and analysis of the notion. Philosophy of science or mathematics at its best usually has a close relation to scientific or mathematical practice. Yet given the divergence in human abilities and the specialisation of our age, it is rare for a single individual to achieve what Pearl has done, that is to innovate both mathematically and philosophically. Indeed the only comparable example I can think of is that of De Finetti, whose works, such as his ([1937]), contain new ideas in both mathematics and philosophy. Pearl's book will therefore be of interest both to mathematicians working in such areas as epidemiology and artificial intelligence, and to philosophers. In this review I will concentrate on the philosophical side of the book, while mentioning briefly some of the mathematical developments.

The present book has developed out of Pearl's earlier ([1988]) book, though with many significant changes. In his ([1988]), Pearl presented a theory of what he called *Bayesian networks*, having introduced the term in his ([1985]). Bayesian networks have proved very successful in artificial intelligence, and are now used in a wide variety of AI applications. To explain the development of Pearl's thought and his present position, it will be necessary to give a brief explanation of this concept. This can most easily be done by giving a simple example, and we have taken the one presented in Pearl ([2000], p. 15), which is illustrated in Figure 1.

The network consists of a finite number of nodes (X_1, X_2, \dots, X_5), some pairs of which are joined by arrows. This network describes a scene that unfolds in a suburb of Los Angeles. The season X_1 (spring, summer, autumn, or winter) influences whether it rains X_2 , or whether a sprinkler is turned on X_3 . Either of these can make the pavement wet X_4 , which in turn can make it slippery X_5 . If an arrow joins X_i to X_j , then X_i is called a *parent* of X_j , and X_j a *child* of X_i . Children, children of children, etc. are called *descendants*. Now

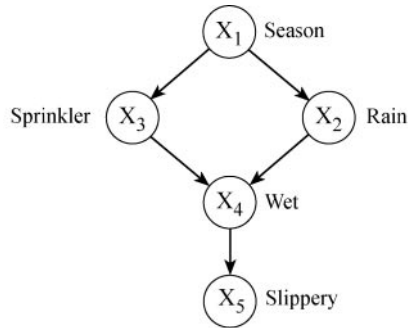


Figure 1

suppose X_1, X_2, \dots, X_5 are random variables with a joint probability distribution. Then the network is called a Bayesian Network if a set of conditional independence relations, known as the *Markov condition*, hold. The Markov condition states that a node is conditionally independent, given its parents, of any set of other nodes that does not contain any of its descendants. The validity of the Markov condition is fairly obvious in the example given. For example, the slipperiness of the pavement is conditionally independent, given its wetness, of the season, whether it is raining, and whether the sprinkler is on.

The network illustrated in Figure 1 is also a *causal network*, that is to say that if an arrow joins X_i to X_j , this can be interpreted as meaning that X_i causes X_j . Although there is a close connection between causal networks and Bayesian networks, it is important to note that they are different concepts. A causal network is a set of causes and effects arranged in a network. One can specify a causal network without mentioning probabilities or random variables. A Bayesian network is a set of random variables with a joint distribution, arranged in a network, and satisfying the Markov condition. One can specify a Bayesian network without mentioning causes. A Bayesian network is a purely probabilistic structure. It is moreover possible to give examples of Bayesian networks that are not causal networks, and causal networks, which are not Bayesian networks, as I will now briefly show.

Suppose we have a Bayesian network that is also a causal network. Assume further, as is usually the case, that the causes are asymmetric, i.e. that, if A causes B, B does not cause A. Now reverse one or more arrows in the network. Because of asymmetry, it ceases to be a causal network. It may also cease to be a Bayesian network, for the Markov condition may now no longer hold. However, we can turn it back into a Bayesian network by the simple expedient of joining with an arrow any two nodes that are conditionally dependent, when, for the Markov condition to hold, they should be

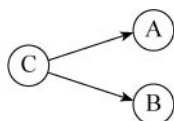


Figure 2

conditionally independent. In this way we generate a Bayesian network that is not a causal network.

To show that causal networks need not be Bayesian networks, I will give an example that is familiar to philosophers of science. In the discussion of causality in his ([1956]), Reichenbach introduced the notion of *conjunctive fork*. Suppose effects A, B have a common cause C. This can be represented in the diagram of Figure 2 by three nodes with an arrow joining C to A, and an arrow joining C to B.

For obvious reasons this is known as a *causal fork*. Reichenbach defined a conjunctive fork to be a causal fork in which A and B were conditionally independent given C. This is just a special case of the Markov condition. So conjunctive forks are Bayesian networks, and the notion of conjunctive form can be considered as a predecessor of that of Bayesian network.

Salmon, investigating the matter further, discovered that not all causal forks are conjunctive forks, and described causal forks, which are not conjunctive forks as *interactive forks*. In his ([1978]) he gives a striking example of an interactive fork—Compton scattering. Let the values of the variables A, B, C in Figure 2 be energies. Let C represent the collision between a high energy photon and a virtually stationary electron. Let A be the electron after the collision, and B the photon after the collision. Suppose $C = E$, $A = E_1$, and $B = E_2$. Then by conservation of energy $E = E_1 + E_2$. So given C, A and B, far from being independent, are highly correlated. It follows that in this case the causal fork is not conjunctive, but interactive. In a different terminology, this is an example of a causal network that is not a Bayesian network.

This gives enough conceptual background to be able to understand the shift in Pearl's thinking from his earlier to his present book. It is clear from what we have said that the relations between Bayesian networks and causal networks are somewhat problematic. Should we lay more emphasis on the probabilistic notion of Bayesian network, perhaps not worrying if not all the arrows in a Bayesian network represent causes? Or should we regard the notion of cause and causal links in the network as fundamental? In a nutshell, Pearl has, since 1988, moved in the direction of regarding causes as more fundamental than probabilities. This is how he puts it ([2000], pp. xiii–xiv):

Ten years ago, when I began writing *Probabilistic Reasoning in Intelligent Systems* (1988), I was working within the empiricist tradition. In this tradition, probabilistic relationships constitute the foundations of human knowledge, whereas causality simply provides useful ways of abbreviating and organizing intricate patterns of probabilistic relationships. Today, my view is quite different. I now take causal relationships to be the fundamental building blocks both of physical reality and of human understanding of that reality, and I regard probabilistic relationships as but the surface phenomena of the causal machinery that underlies and propels our understanding of the world.

Probability has, however, had an advantage over causality in that it has been mathematized since the eighteenth century, whereas cause has, until very recently, been a qualitative notion for which no mathematical calculus existed. In order to bring cause front stage, therefore, Pearl has undertaken the task of mathematizing causality. This has led to the many interesting mathematical developments with which his present book is filled. I will now say something very brief about these before turning to the philosophical questions.

Pearl's mathematization of causality can conveniently be divided into three aspects. The first of these is the use of network diagrams. Even our simple example given in Figure 1 shows how useful these can be. The arrows between nodes are a natural representation of causes, since these are, for the most part, asymmetrical. Moreover, a cause can have several effects, and an effect several causes. A causal network represents all this complicated information in a single diagram. Pearl, quite rightly in my view, stresses the importance of diagrams for mathematical reasoning. He gives an example of a circuit diagram ([2000], p. 345), and comments ([2000], p. 344):

This diagram is, in fact, one of the greatest marvels of science. It is capable of conveying more information than millions of algebraic equations or probability functions or logical expressions. What makes this diagram so much more powerful is the ability to predict not merely how the circuit behaves under normal conditions but also how the circuit will behave under millions of *abnormal* conditions.

Causal networks or graphs may then play a crucial role in giving causality a central place in science. As Pearl puts it ([2000], p. 138):

Recent developments in graphical methods promise to bring causality back into the mainstream of scientific modeling and analysis [...] the crucial change has been the emergence of graphs as a mathematical language. This mathematical language is not simply a heuristic mnemonic device for displaying algebraic relationships [...] Rather, graphs provide a fundamental notational system for concepts and relationships that are not easily expressed in the standard mathematical languages of algebraic equations and probability calculus.

I don't think anyone could seriously doubt the value of causal diagrams, and that their development has been a considerable step forward in handling causality.

The second aspect of Pearl's mathematization of causality is his introduction of what he calls the *do* operator. This can again be illustrated by the simple example of a causal network illustrated in Figure 1. Let us suppose first that I am walking along a street in Los Angeles and observe that a sprinkler is on in a front garden. Such an *observation* can be distinguished from an *intervention* in which I am living in the house in question, go into the garden and turn the sprinkler on. Pearl proposes representing such interventions by means of his *do* operator as $do(X_3 = \text{On})$. This also involves a surgery on the network that consists in removing the arrow joining X_1 to X_3 .

Pearl applies his *do* operator to the classic problem of smoking and lung cancer. It was found from statistics that getting lung cancer is correlated with smoking, or in symbols

$$P(\text{lung cancer} \mid \text{smoking}) > P(\text{lung cancer}) \quad (1)$$

But does this correlation indicate a causal link? The tobacco companies argued that perhaps it did not. It could be that a particular genetic disposition inclined some people both to smoke and to contract lung cancer. The cause of the lung cancer would then be this genetic disposition, and not smoking. People who did not have the genetic disposition in question could smoke as much as they pleased without any increased chance of getting lung cancer. This hypothesis of the tobacco companies can be expressed using Pearl's *do* operator as

$$P(\text{lung cancer} \mid do(\text{smoking})) = P(\text{lung cancer}) \quad (2)$$

Now, we could in theory test whether (2) was true by carrying out a randomized trial, but this is obviously not possible for ethical and social reasons. Pearl carries out a mathematical development that could help to overcome this problem.

He first of all works out an algebra of the *do* operator. Let us now consider a particular problem in which the *do* operator is involved, but in which we cannot perform controlled experiments. We could still get results of evidential bearing on the question if we could eliminate occurrences of the *do* operator using this algebra. Diagrams play an important role in this process since they are used to guide the surgery on equations that is needed by the algebra. All this is most ingenious, and Pearl's algebra of the *do* operator will no doubt find many practical applications—perhaps particularly in the field of epidemiology, where it is especially important to distinguish causal correlations from non-causal correlations.

The third aspect of Pearl's mathematization of causality is his use of *functional causal models*. He explains this concept as follows ([2000], p. 27):

a functional causal model consists of a set of equations of the form

$$x_i = f_i(pa_i, u_i), i = 1, \dots, n,$$

where pa_i (connoting *parents*) stands for the set of variables judged to be immediate causes of X_i and where U_i represent errors (or 'disturbances') due to omitted factors.

If the graph corresponding to a functional causal model is acyclic, and the error terms are mutually independent, the model is called *Markovian*. Pearl shows ([2000], p. 30) that a Markovian causal model is a Bayesian network.

Pearl's functional causal models were developed from the structural equation models used in econometrics. Curiously, however, as Pearl points out ([2000], p. 137), econometricians have recently moved away from interpreting their structural equations causally, preferring to give them a purely probabilistic sense. Pearl speculates that this may be because of the lack of adequate mathematical formalism for handling causality, and hopes that his own mathematization of causality may reverse the trend.

This brings us to the central question posed by Pearl's new book. He has already persuaded research workers in AI and related fields to adopt his concept of Bayesian network. Should such workers stick to the essentially probabilistic concept of Bayesian network or change to Pearl's new notion of functional causal model that is essentially causal rather than probabilistic in character? Naturally this general question will come to be decided by the success or otherwise of the new mathematical techniques in various applications. However there is a philosophical side to this problem, and, as this is the most fundamental philosophical question posed by Pearl's book, I will devote the rest of this review to considering it.

The philosophical issue concerns the relative importance of causality and probability. On this question we can distinguish three positions. (1) *The no causality view*. This is the view that cause is an arcane and unnecessary notion that should be eliminated as far as possible from science. This position appears to have been popular in Edwardian England, since it is expressed by Karl Pearson in the 1911 third edition of his *The Grammar of Science* (quoted by Pearl [2000], p. 340), and by Bertrand Russell in his ([1913]) article, 'On the Notion of Cause' (quoted by Pearl [2000], p. 337). (2) *The weak causality view*. This is the view that causality though an important concept for science is less suited than probability for mathematical theories. It is a desirable, even perhaps necessary, adjunct for such theories, but remains ancillary and qualitative. This might be the view of someone who preferred to use Bayesian networks in constructing mathematical models, but still thought it desirable that some, though perhaps not all, arrows in such networks should have a

causal interpretation. (3) *The strong causality view*. This is the view that causality is a more important and fundamental notion than probability, and should be used in preference to probability as the core of mathematical models. This, I think, is Pearl's view in the present book, as is shown by the quotation from the preface given earlier, and also by his emphasis on functional causal models. It is worth noting that in such models probability only appears in connection with the error terms U_i . What I want to argue next is that Pearl makes a decisive case against the 'no causality view', but that the 'weak causality view' might perhaps be defended against his 'strong causality view'.

Perhaps Pearl's strongest argument against the no causality view of Pearson and Russell is his admirable treatment of Simpson's paradox that occurs in Section 6.1 (pp. 174–82) of his book. Suppose we have a population that consists of the union of a set of disjoint sub-populations. It can happen that C increases the probability of E in the whole population, while at the same time decreasing the probability of E in every one of the set of sub-populations. Many people have regarded this reversal as paradoxical, and it has come to be known as Simpson's paradox. The most famous case of the paradox occurred in connection with admissions to Berkeley's graduate school. It was discovered that overall male applicants had a higher probability of getting admitted than female applicants, which naturally suggested that there was gender bias in the graduate admissions. However, further analysis revealed that female applicants had a higher probability of gaining admission to every single department. Pearl emphasizes that it is impossible to explain what is going on here if we follow the recommendation of Pearson and Russell to abstain from using causality, and hence confine ourselves to probability statements. Once we introduce causality considerations, however, the matter is simply explained. The cause of the lower overall probability of female applicants gaining admission was not gender bias as at first appeared to be the case, but simply the fact that female applicants had a greater tendency to apply to popular departments where the rejection rate was higher. Their lower overall probability of gaining admission was thus perfectly compatible with having a slightly higher probability of gaining admission to each individual department. A proper explanation of why such reversals appear to be paradoxical does therefore require the use of causal concepts, but it only requires the use of such concepts in an ancillary, qualitative role. This case does not therefore decide the issue between the weak causality view and the strong causality view.

Pearl has another simpler but yet decisive argument against the no causality view. It goes as follows ([2000], p. 134):

It is an embarrassing yet inescapable fact that probability theory [...] does not permit us to express sentences such as 'Mud does not cause rain'; all we can say is that the two events are mutually correlated, or

dependent—meaning that if we find one we can expect to encounter the other.

From the point of view of the weak causality position, however, it could be claimed that statements like ‘Mud does not cause rain’ might have an informal ancillary character, and not be part of the mathematical calculus. I will now consider how the weak causality view might be defended against the strong causality view.

In order to make such a defence, it is necessary to change one feature of Pearl’s position. Pearl (again like De Finetti!) always adopts a degree of belief interpretation of probability. Thus in his earlier book he says ([1988], p. 29):

Bayesian methods provide a formalism for reasoning about partial beliefs under conditions of uncertainty. In this formalism, propositions are given numerical parameters signifying the degree of belief accorded them under some body of knowledge, and the parameters are combined and manipulated according to the rules of probability theory.

Pearl continues to hold the same point of view in the present book, writing ([2000], p. 2):

We will adhere to the Bayesian interpretation of probability, according to which probabilities encode degrees of belief about events in the world and data are used to strengthen, update, or weaken those degrees of belief.

A Bayesian network, however, is defined in terms of probabilities, and it is often possible to interpret these probabilities as objective rather than subjective probabilities, since there are two well-established objective interpretations of probability—the frequency and the propensity. Consider Pearl’s own example of a Bayesian network illustrated in Figure 1. If we take as our underlying population houses with gardens in a Los Angeles suburb, then by taking a large enough random sample of such houses at different seasons of the year, we could collect statistical data from which it would be possible to infer all the probability distributions in the Bayesian network entirely objectively without considering the degrees of belief of any individual. Moreover, we could check from this data whether the Markov condition really held. If probability is interpreted in this objective fashion, then a case can be made for giving probability the leading role in mathematical models, and considering causality as a necessary but ancillary and qualitative concept. To see this let us consider some of Pearl’s arguments for the opposite view, i.e. the strong causality view.

Pearl writes ([2000], p. 25):

causal relationships are more ‘stable’ than probabilistic relationships. We expect such difference in stability because causal relationships are *ontological*, describing objective physical constraints in our world,

whereas probabilistic relationships are *epistemic*, reflecting what we know or believe about the world.

I agree with Pearl that causes are objective, but would deny that probabilities need be just epistemic. They too can be objective, and describe physical features of the world. Moreover, I would argue that probabilistic features of the world are often just as stable as causal features, and sometimes more so.

Pearl gives as an example of a stable causal relationship: 'Turning the sprinkler on would not affect the rain', and remarks that this ([2000], p. 25) 'remains true regardless of what we learn or know about the season or about the pavement.' He is certainly right about this example, but the same does not apply to all causal relationships. Consider, for example: 'Turning this handle will cause water to come out of the sprinkler.' This is hardly a very stable causal relationship. I would doubt its truth if I learn that the supply of water to the neighbourhood has been temporarily cut off, or that the mechanism joining the handle to the valve has been disconnected to prevent the children playing with the sprinkler and getting themselves soaked. Conversely there are many very stable probabilistic relationships. For example, the radioactive emission of alpha particles follows a Poisson distribution, the size of adult animals in a certain population follows a normal distribution, the distribution of incomes in a capitalist society follows a lognormal distribution for low incomes, and a Pareto distribution for moderate and high incomes. All these probabilistic laws, and there are many like them, are stable and objective features of the external world.

What I have called Pearl's strong causality view is clearly shown in its support for a Laplacian determinism. He writes ([2000], p. 26):

In this book, we shall express preference toward Laplace's quasi-deterministic conception of causality and will use it, often contrasted with the stochastic conception, to define and analyse most of the causal entities that we study [...] the Laplacian conception is more in tune with human intuition. The few esoteric quantum mechanical experiments that conflict with the predictions of the Laplacian conception evoke surprise and disbelief, and they demand that physicists give up deeply entrenched intuitions about locality and causality [...] Our objective is to preserve, explicate, and satisfy—not destroy—those intuitions.

Laplace's views on determinism were formulated at a time when it was generally believed that Newtonian mechanics applied to all particles from the smallest to the largest. Quantum mechanics has shown that this assumption is false. Pearl seems to think that this is not relevant except to the world of micro particles where quantum mechanics applies. But many phenomena in the macro world may well be amplifications of micro events, and consequently an indeterministic view of the macro world seems to be more in accord with modern science.

I would also criticize Pearl's reliance on human intuition. History shows that human intuition can often lead humans astray, and is frequently based on ideas that can be superseded by the advance of science. For many centuries it was obvious to human intuition that the earth was flat, that the sun moved, etc. The work of Tversky and Kahneman has shown that probabilistic reasoning is not very intuitive for humans. However, when human intuition gives one answer, and the probability calculus another, we usually say that human intuition is wrong, not that the probability calculus needs to be altered. A calculus which has proved successful in a wide variety of applications, and which allows precise mathematical reasoning, is rightly preferred to highly fallible human intuitions. Pearl does quote one of Tversky and Kahneman's results ([2000], p. 22). Their finding was that most people believe that it is more likely for a girl to have blue eyes, given that her mother has blue eyes, than the other way around, whereas the two probabilities are in fact equal. Pearl correctly takes this as supporting the claim that (2000, p. 22): 'people tend to ignore probabilistic information altogether and attend to causal information instead.' What is questionable, however, is whether people are always right to prefer causal to probabilistic information. In this instance the strategy led them to the wrong answer. There may indeed be other cases in which it is a good strategy to replace intuitive causal reasoning by reasoning based on the mathematical theory of probability, despite the fact that the use of causality is often indispensable.

References

- De Finetti, B. [1937]: 'Foresight: Its Logical Laws, Its Subjective Sources', English trans. in H. E. Kyburg and H. E. Smokler (eds) [1964], *Studies in Subjective Probability*, Wiley, pp. 93–158.
- Pearl, J. [1985]: 'Bayesian Networks: a Model of Self-activated Memory for Evidential Reasoning', in *Proceedings, Cognitive Science Society*, Ablex, pp. 329–34.
- Pearl, J. [1988]: *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann.
- Pearl, J. [2000]: *Causality. Models, Reasoning, and Inference*, Cambridge University Press.
- Pearson, K. [1911]: *The Grammar of Science*, third edition, Meridian Books [1957].
- Reichenbach, H. [1956]: *The Direction of Time*, University of California Press.
- Russell, B. [1913]: 'On the Notion of Cause', reprinted in *Mysticism and Logic and Other Essays* [1917], George Allen & Unwin, pp. 132–51.
- Salmon, W. C. [1978]: 'Why Ask, "Why?"? An Inquiry Concerning Scientific Explanation', in W. C. Salmon [1998], *Causality and Explanation*, Oxford University Press, Ch. 8, pp. 125–41.